# Region-Growing Fully Convolutional Networks for Hyperspectral Image Classification with Point-Level Supervision

Yonghao Xu[1], Pedram Ghamisi[1,2]

[1]Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria
[2]Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, 09599 Freiberg, Germany

**Abstract**
Deep learning algorithms have shown their great potential in the hyperspectral image (HSI) classification task. However, training these models generally requires a large number of training samples, which are laborious and time-consuming to collect in practice. To reduce the annotation burden, we propose to conduct HSI classification with point-level supervision, where only one annotated pixel in each category would be used for training. To this end, a novel region-growing fully convolutional network (RGFCN) is proposed in this study. The key idea of RGFCN is to expand the annotated regions from the original one point in each category with the region growing technique. As the annotated region grows, the network could also get stronger supervision, which in turn, helps to produce more accurate pseudo labels in the region growing. To better train the proposed RGFCN, we further adopt the entropy minimization strategy to assist the training in those unlabeled regions. Experiments on two benchmark HSI datasets demonstrate the effectiveness of the proposed approach.

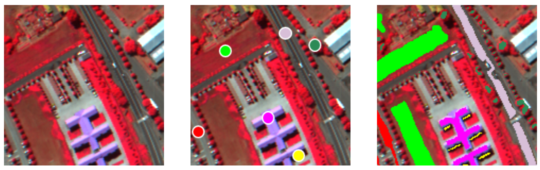
**Keywords**
Weakly supervised classification, hyperspectral image (HSI), convolutional neural networks, deep learning

## 1. Introduction

As an important data source for Earth Observation, hyperspectral images (HSIs) can record detailed information from the Earth's surface in both spatial and spectral domains with hundreds or even thousands of continual spectral bands. Because of this property, HSIs have been widely used in many applications, such as environment monitoring, urban planning, and resource exploration [1]. To achieve these applications, HSI classification is the fundamental task that aims to assign a class label for each pixel in the image.

Witnessing the great development of deep learning algorithms in the computer vision field, recent research has attempted to use advanced deep neural networks to tackle the HSI classification task [2]. However, considering the high complexity of deep neural networks, training these models generally requires a large number of accurate pixel-wise annotations, which are very laborious and time-consuming to collect in practice [3].

The main burden of assigning pixel-wise labels for a remote sensing image comes from the boundary regions of different objects. Figure 1 presents an example. It can be observed that annotating the detailed boundaries for each object in the HSI is difficult since the spatial distribution of the Earth's surface is very complex. To reduce the annotation burden, this paper proposes to conduct HSI classification with point-level supervision, where only one pixel in each category is annotated with the class label. As shown in Figure 1 (b), the collection of point-level annotation is much easier for an annotator expert.



(a) (b) (c)

**Figure 1:** An illustration of different types of annotations for HSI classification. (a) The false color image. (b) Point-level annotation used in this work. There is only one pixel in each category annotated with ground-truth label. (c) Pixel-wise annotation used in previous study.

Compared to the traditional classification scenario where the machine learning models could obtain sufficient training data, classification with point-level supervision is much more challenging especially for supervised deep learning models which naturally require a large training set. Besides, the complex spatial distribution and the spectral heterogeneity of objects in HSIs

CDCEO 2021: 1st Workshop on Complex Data Challenges in Earth Observation, November 1, 2021, Virtual Event, QLD, Australia.
yonghaoxu@ieee.org (Y. Xu); pedram.ghamisi@iarai.ac.at (P. Ghamisi)
0000-0002-6857-0152 (Y. Xu); 0000-0003-1203-741X (P. Ghamisi)

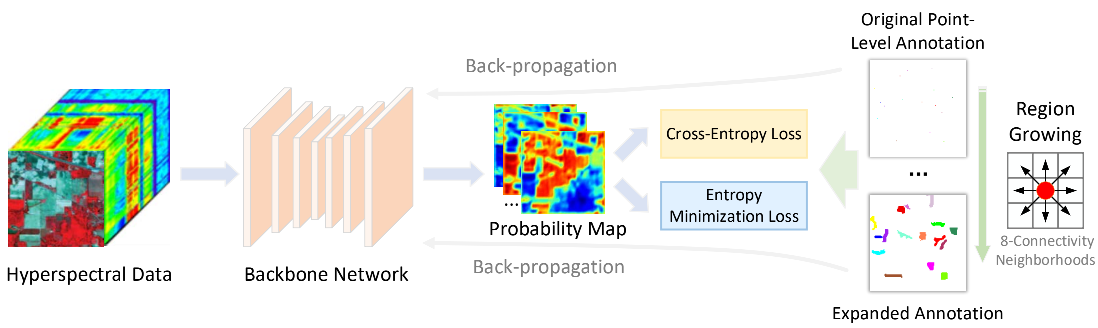



**Figure 2:** An illustration of the proposed region-growing fully convolutional networks.

make it more difficult to achieve satisfactory classification performance with weak supervision. To tackle the aforementioned challenges, we propose a region-growing fully convolutional network (RGFCN) for HSI classification with point-level supervision. The initial inspiration of this work comes from an observation that adjacent pixels in remote sensing images, in particular those that are of high spatial resolution, tend to belong to the same category considering the spatial continuity of objects. To this end, we propose a novel region-growing mechanism that helps to expand the annotated regions from the original one point in each category. As the annotated region grows, the network could also get stronger supervision, which in turn, helps to produce more accurate pseudo labels in the region growing. To better train the proposed RGFCN, we further adopt the entropy minimization strategy to assist the training in those unlabeled regions. Experiments on two benchmark HSI datasets demonstrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 describes the proposed RGFCN in detail. Section 3 presents the experiments in this study. Conclusions and other discussions would be made in Section 4.

# 2. Region-Growing Fully Convolutional Networks

One of the main challenges of HSI classification with point-level supervision comes from the insufficiency of training samples [4]. To alleviate this problem, we propose to use the region growing mechanism to gradually expand the training samples from the original one pixel in each category. The initial inspiration of this idea comes from an observation that neighboring pixels tend to belong to the same category considering the spatial continuity of ground objects [5]. Thus, a natural idea is to conduct region growing and network training simultaneously. As the annotated training samples expand, the network could get stronger supervision, which in turn, would also help to generate more accurate pseudo labels in the region growing.

Based on the aforementioned idea, we propose a region-growing fully convolutional network (RGFCN), as shown in Figure 2. In what follows, we will describe the proposed region-growing mechanism and the optimization of the network in detail.

## 2.1. Region-Growing Mechanism

Formally, let $X \in \mathbb{R}^{h\times w\times n}$ be the input HSI, where $h$, $w$, and $n$ are the height, width, and the number of bands in the image, respectively. Let $F(\cdot)$ denote the mapping function of the backbone network, and $P = F(X) \in \mathbb{R}^{h\times w\times k}$ be the corresponding probability map of $X$, where $k$ is the number of total categories. Recall that our goal is to expand the annotated regions iteratively. To this end, at each iteration in the training phase, we visit all labeled pixels $x_l \in L$, where $L$ denotes the set that collects the locations of labeled pixels at the current iteration. Specifically, for each $x_l \in L$, let $p \in [1, k]$ be its label and $C_8$ denote the 8-connectivity neighborhood regions of $x_l$. Then, for each unlabeled pixel $x_{(u,v)} \in C_8$, the following criterion is used to update its label:

$$The\,label\,of\;x_{(u,v)} \leftarrow p,\;if\;\begin{cases}\mathrm{argmax}\left(P_{(u,v,p\prime)}\right) = p\\ P_{(u,v,p)} \geq \tau,\end{cases} \tag{1}$$

where $P_{(u,v,p\prime)}$ denotes the probability of the $p\prime$th class at position $(u, v)$, and $\tau$ is a confidence threshold.

In this way, the annotated regions could gradually expand from the original one pixel in each category.

## 2.2. Optimization

Obviously, the quality of the expanded annotation depends largely on the accuracy of the probability map $P$

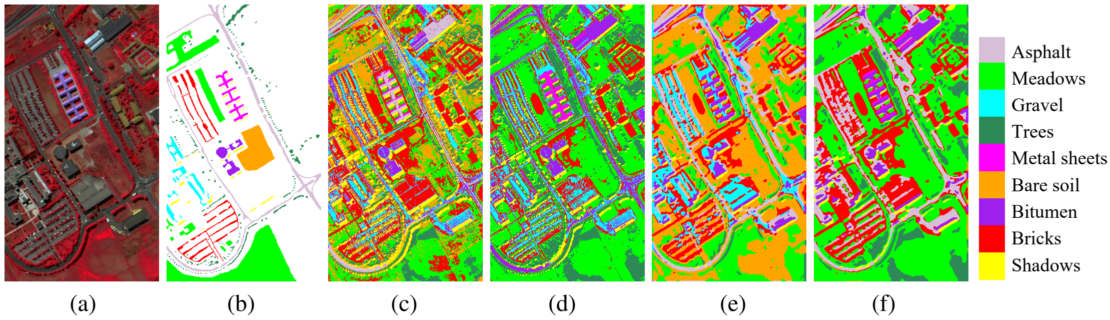



**Figure 3:** Classification maps for the Pavia University dataset with point-level supervision. (a) The false color image. (b) Ground-truth map. (c) SpeFCN only. (d) SpeFCN-RGFCN. (e) SpaFCN only. (f) SpaFCN-RGFCN.

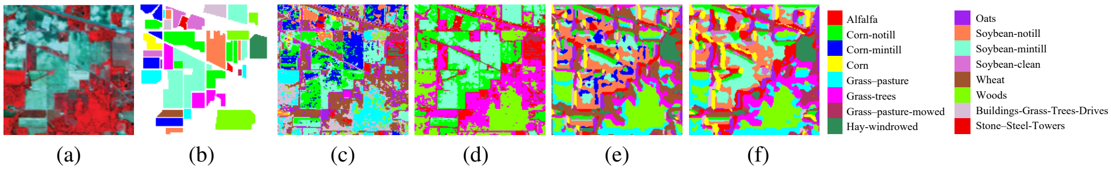



**Figure 4:** Classification maps for the Indian Pines dataset with point-level supervision. (a) The false color image. (b) Ground-truth map. (c) SpeFCN only. (d) SpeFCN-RGFCN. (e) SpaFCN only. (f) SpaFCN-RGFCN.

in (1). To constrain the optimization of the network, we first define the cross-entropy loss $\mathcal{L}_{cls}$ on the labeled set $L$ as:

$$\mathcal{L}_{cls} = -\frac{1}{|L|}\sum_{(u,v)\in L}\sum_{i=1}^{k} Y_{(u,v,i)} \log\left(P_{(u,v,i)}\right), \quad (2)$$

where $Y$ is the label map at current iteration.

Note that the annotations in $Y$ would be very sparse in practice since there are a lot of unlabeled samples. To better train the network and apply constraint on those unlabeled regions, the entropy minimization strategy is adopted [6]. In fact, the generated probability map $P$ behaves as a discrete distribution over different classes. If the network has very low confidence in recognizing the class label in position $(u, v)$, its probabilities $P_{(u,v,i)}$ would be evenly spread in different classes. In this case, the entropy at position $(u, v)$ would also be large. Thus, to encourage the network make more confident predictions on those unlabeled regions, the entropy minimization loss $\mathcal{L}_{ent}$ can be defined as:

$$\mathcal{L}_{ent} = -\frac{1}{|U|\log(k)}\sum_{(u,v)\in U}\sum_{i=1}^{k} P_{(u,v,i)} \log\left(P_{(u,v,i)}\right), \quad (3)$$

where $U$ denotes the set that collects the locations of current unlabeled pixels. With the constraint in (3), the network would be more likely to produces high-confident predictions.

The final optimization of the network can be formulated as:

$$\min_{\theta} \mathcal{L}_{cls} + \lambda_{ent}\mathcal{L}_{ent}, \quad (4)$$

where $\theta$ denotes the parameters in the backbone network, and $\lambda_{ent}$ is a weighting factor for the entropy minimization loss.

Note that the optimization of the network and the region growing procedure are simultaneous. At each iteration, we first use the back-propagation algorithm to update the parameters $\theta$. Then, the labeled regions $L$ and the unlabeled regions $U$ get updated through the region growing mechanism.

# 3. Experiments

## 3.1. Data Description

The Pavia University dataset[1], and the Indian Pines dataset[1] are utilized to evaluate the performance of the proposed method. For each category in both datasets, we randomly select 1 labeled pixel in each category from the ground truth data to make up the original training set, while the remaining samples are used as the test set.

[1] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

**Table 1**
Quantitative results of the Pavia University dataset.

| Backbone | Model | OA (%) | $\kappa$ (%) | AA (%) |
|---|---|---|---|---|
| SpeFCN | Backbone Only | 48.77 | 38.24 | 57.02 |
| | RGFCN | 55.16 | 42.55 | 62.48 |
| SpaFCN | Backbone Only | 45.17 | 34.29 | 48.21 |
| | RGFCN | 62.27 | 49.49 | 51.45 |

**Table 2**
Quantitative results of the Indian Pines dataset.

| Backbone | Model | OA (%) | $\kappa$ (%) | AA (%) |
|---|---|---|---|---|
| SpeFCN | Backbone Only | 27.01 | 18.61 | 42.21 |
| | RGFCN | 41.05 | 30.72 | 35.35 |
| SpaFCN | Backbone Only | 30.56 | 23.43 | 38.10 |
| | RGFCN | 37.98 | 31.22 | 51.92 |

### 3.2. Performance Evaluation

In this subsection, we report the classification results of the proposed RGFCN. Two different backbone networks including the SpeFCN [7], and SpaFCN [7] are adopted to implement RGFCN, respectively. The confidence threshold $\tau$, and the weighting factor $\lambda_{ent}$ are set as $0.3$ and $0.01$ in the experiments, respectively. The overall accuracy (OA), kappa coefficient ($\kappa$), and average accuracy (AA) are utilized to quantitatively estimate different methods. We implement the experiments using the PyTorch platform with an Intel Xeon E5-2678 2.50-GHz CPU and one NVIDIA GeForce RTX 2080 Ti GPU.

As shown in Table 1 and Table 2, the existing state-of-the-art deep neural networks like SpeFCN and SpaFCN can hardly yield good performance in both datasets since there is only one labeled pixel for each category in the training set. By contrast, with the help of the proposed RGFCN, the performance gets dramatically improved. Take the result of the SpeFCN on the Indian Pines dataset for example. While the backbone network can only yield an OA of about 27%, the OA of RGFCN can reach more than $41\%$, which outperforms the previous one with about $14\%$. Similar phenomenons can be observed in other scenarios.

To visually evaluate the classification results, we further present the classification maps in Figure 3 and Figure 4. It can be observed that there exist a lot of salt and pepper noises in the results of SpeFCN and SpaFCN due to the insufficiency of training samples. By contrast, RGFCN can significantly improve the quality of the classification maps especially for objects with large spatial sizes like meadows.

## 4. Conclusions and Discussions

This paper proposes a region-growing fully convolutional network (RGFCN) for HSI classification with point-level supervision. To tackle the challenge of insufficient training samples, a novel region growing mechanism is proposed. Besides, the entropy minimization loss is adopted to further constrain the training on those unlabeled regions. Experiments on two benchmark HSI datasets demonstrate the effectiveness of the proposed method.

Since the performance of the whole framework depends largely on the quality of the region growing mechanism, once wrong annotations are included in the expanded labeled regions, the network may be misguided to make wrong predictions. Thus, we will try to further improve the region growing strategy to better filter out those unreliable samples in our future work.

## References


[1] P. Ghamisi, J. Plaza, Y. Chen, J. Li, A. J. Plaza, Advanced spectral classifiers for hyperspectral images: A review, IEEE Geosci. Remote Sens. Mag. 5 (2017) 8–32.

[2] Y. Xu, L. Zhang, B. Du, F. Zhang, Spectral-spatial unified networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 56 (2018) 5893–5909.

[3] Y. Xu, B. Du, L. Zhang, D. Cerra, M. Pato, E. Carmona, S. Prasad, N. Yokoya, R. Hänsch, B. Le Saux, Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 ieee grss data fusion contest, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 12 (2019) 1709–1724.

[4] L. Zhang, L. Zhang, B. Du, Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geosci. Remote Sens. Mag. 4 (2016) 22–40.

[5] R. Adams, L. Bischof, Seeded region growing, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1994) 641–647.

[6] T.-H. Vu, H. Jain, M. Bucher, M. Cord, P. Pérez, Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 2517–2526.

[7] Y. Xu, B. Du, L. Zhang, Beyond the patchwise classification: Spectral-spatial fully convolutional networks for hyperspectral image classification, IEEE Trans. Big Data. 6 (2019) 492–506.